

**PRIORITY
DOCUMENT**

SUBMITTED OR TRANSMITTED IN
COMPLIANCE WITH RULE 17.1(a) OR (b)



REC'D 14 SEP 2004

WIPO PCT

**Prioritätsbescheinigung über die Einreichung
einer Patentanmeldung**

Aktenzeichen: 103 48 408.6

Anmeldetag: 14. Oktober 2003

Anmelder/Inhaber: DaimlerChrysler AG, 70567 Stuttgart/DE

Bezeichnung: Nutzeradaptive Dialogunterstützung für Sprachdialogsysteme

IPC: G 10 L 11/02

**Die angehefteten Stücke sind eine richtige und genaue Wiedergabe der
ursprünglichen Unterlagen dieser Patentanmeldung.**

München, den 13. Juli 2004
Deutsches Patent- und Markenamt
Der Präsident
Im Auftrag

Stanschus

DaimlerChrysler AG

Finkele

Nutzeradaptive Dialogunterstützung für Sprachdialogsysteme

Die Erfindung betrifft ein Verfahren zur nutzeradaptiven Dialogunterstützung bei Sprachdialogsystemen nach dem Oberbegriff des Patentanspruchs 1.

Zur Bedienung komplexer technischer Einrichtungen, insbesondere von Assistenzsystemen bei Kraftfahrzeugen, werden zunehmend Sprachdialogsysteme (Spracherkennungssysteme) eingesetzt, da hierbei angenommen wird, dass eine rein sprachliche Interaktion den Bediener der technischen Einrichtung weniger von seiner primären Bedienungsaufgabe ablenkt, als dies eine haptisch-visuelle Bedienung bedingen würde.

Bei Sprachdialogsystemen besteht jedoch generell das Problem, dass man mit dem System unterschiedlich erfahrene Nutzer sprachlich möglichst optimal bedienen muss; beispielsweise einen Anfänger, der mit dem System nicht vertraut ist, oder aber einen Experten, welcher das System in allen Einzelheiten und Feinheiten kennt und beherrscht. Entsprechend dieser unterschiedlichen Vertrautheit mit dem System werden andere Anforderungen an die Bedienung des Sprachdialogsystems gestellt. Der Anfänger benötigt mehr Hilfestellung und Führung durch das System, um dieses auf dem Wege des learning-by-doing kennen lernen zu können. Der Experte jedoch möchte eine möglichst schnelle und effektive Interaktion mit dem Sprachdialogsystem. Darüber hinaus werden moderne Sprachdialogsysteme immer komplexer, da die Vielfalt der zu bedienenden Funktionen steigt. Dies impliziert, dass es in Zukunft nicht mehr

den Experten oder den Anfänger geben wird. Es wird Benutzer geben, die einen Teil der angebotenen Funktionalitäten häufig bedient und die in diesem Teil Experte sind, und es wird Benutzer geben, die sich wiederum nur in einem anderen Teil des Systems auskennen.

Es existieren Sprachdialogsysteme bei welchen es dem Systemnutzer möglich ist anzugeben, wie gut er mit dem System bereits vertraut ist. Dementsprechend interagiert das Dialogsystem mit dem Systemnutzer über kürzere oder längere Systemäußerungen (Sprachaufforderungen). Die Einstellen in Bezug auf den Vertrautheitsgrad sind jedoch durch den Systemnutzer aktiv einzugeben und die jeweiligen Einstellungen beziehen sich sodann auf den gesamten Dialog. Dies deckt somit nicht diejenigen Fälle ab, bei denen ein Systemnutzer sich beispielsweise an sich recht gut mit dem Sprachdialogsystem auskennt, aber bei einem Dialogschritt vergessen hat, welche Äußerung auf eine Sprachaufforderung durch das System erwartet wird, um im Dialog sinnvoll voranzuschreiten. Hier hilft es dem Systemnutzer nicht, dass er die Möglichkeit hat die Systemeinstellung in Bezug auf seinen Vertrautheitsgrad zu ändern und hierdurch mitzuteilen, dass er mehr Unterstützung durch das Sprachdialogsystem benötigt, da in den nachfolgenden Dialogschritten diese Unterstützung wiederum nicht mehr benötigt wird. Problematisch ist hierbei zudem, dass durch die erforderliche Eingabe des Vertrautheitsgrades die Systemfunktionalität stark von der Selbsteinschätzung des Systemnutzers abhängt.

Es ist deshalb wünschenswert, dass das Sprachdialogsystem im Fall, dass der Systemnutzer Schwierigkeiten bei der Eingabe der notwendigen Sprachäußerungen hat, automatisch Unterstützung anbietet. Ein derartiges System wird in der Offenlegungsschrift US 2002/0147593 A1 beschrieben. Hierbei ist das Sprachdialogsystem in der Lage zwei unterschiedlich detaillierte Sprachaufforderung auszugeben, jeweils in Abhängigkeit

davon ob das System davon ausgeht ob es sich bei dem Systemnutzer um einen unterstützungsbedürftigen Anfänger oder um einen versierten Experten handelt. Bei der Kommunikation mit einem Anfänger verwendet das Sprachdialogsystem Sprachaufforderungen (Prompt) in der für solche Systeme üblichen Detaillierungsgrad, gibt also ausreichend Hinweise auf die Art und Weise der im Rahmen des Dialoges sinnvoller Weise erwarteten Nutzeräußerung. Handelt es sich bei dem Systemnutzer um einen Experten, so wird nur eine verkürzte, optimierte Sprachaufforderung („tapered“ Prompt) ausgegeben. In der Regel enthalten diese verkürzten Sprachaufforderungen keine oder nur sehr wendige erklärende oder unterstützende Hinweise. Während des Dialogverlaufes schätzt das Sprachdialogsystem kontinuierlich den Systemnutzer bezüglich seines Erfahrungsgrades ein und gestaltet entsprechend seine Sprachaufforderungen. Da das System bei der Initiierung des Sprachdialoges nichts über den Systemnutzer weiß, werden zuerst Sprachaufforderungen mit dem üblichen Detaillierungsgrad. In denjenigen Fällen, in welchem im Laufe des Dialoges festgestellt wird, dass der Systemnutzer über eine gewisse Anzahl von aufeinander folgenden Dialogschritten sinnvoll auf die Sprachforderungen reagiert, wird davon ausgegangen, dass es sich hierbei um einen Experten handelt, worauf die dieser Einschätzung nachfolgenden Sprachaufforderungen in Form eines Short-Prompts erfolgen. Da diese Einschätzung jedoch fehlerhaft sein kann, wird die Ausgabe von Short-Prompts nur solange fortgesetzt, solange der Systemnutzer auch fehlerfrei und sinnvoll auf diese reagiert. Reagiert der Systemnutzer auf die Short-Prompts mit Äußerungen, welche das Sprachdialogsystem nicht sinnvoll weiter verarbeiten kann, geht dieses dazu über bei der wiederholten Anfrage und nachfolgend wieder Sprachaufforderungen mit dem üblichen Detaillierungsgrad zu generieren. Eine Rückkehr zur Verwendung der Short-Prompts erfolgt erst wieder nachdem wiederum bei einer gewissen Anzahl von aufeinander folgenden Di-

alogschritten sinnvoll auf die detaillierten Sprachaufforderungen reagiert wurde. Dieses zurückschalten den für den unerfahrenen Systemnutzer gedachten detaillierten Sprachaufforderungen ist notwendig, da das Sprachdialogsystem einzig auf Grund der Art und Weise der Äußerung auf die Sprachaufforderung auf den Erfahrungsgrad des Systemnutzers schließen kann. Problematisch ist hierbei, dass in den Fällen, in welchen ein Experte beispielsweise auf Grund einer Ablenkung eine Fehleingabe vornimmt, dieser nachfolgend wiederholt und unnötig detaillierte Sprachaufforderungen erhält, durch welche er sich gestört fühlen könnte.

Aufgabe der Erfindung ist es deshalb, eine nutzeradaptive Dialogführung für Sprachdialogsysteme zu finden, welche unerfahrene und erfahrene Systemnutzer unterscheidet, und daran angepasste Sprachaufforderungen so generiert, dass auch in denjenigen Fällen, in welchen ein erfahrender Nutzer innerhalb eines Dialogschrittes fehlerhaft reagiert hat, er in den nachfolgenden Schritten, ohne Nachteil für unerfahrene Nutzer, umgehend wieder wie ein erfahrener Nutzer behandelt wird.

Die Aufgabe wird durch ein Verfahren mit den Merkmalen des Patentanspruchs 1 gelöst. Vorteilhafte Ausgestaltungen und Weiterbildungen der Erfindung sind durch die Unteransprüche beschrieben.

Bei dem Verfahren für eine nutzeradaptive Dialogführung gibt ein Sprachdialogsystem eine Sprachaufforderung aus, wobei hierauf das Sprachdialogsystem auf eine Äußerung durch den Systemnutzer wartet. Hierbei wird, um die Nutzeräußerung zu verstehen, ein Spracherkenner aktiviert wird. Das Sprachdialogsystem ist in der Lage unerfahrene und erfahrene Nutzer zu unterscheiden, wobei es unerfahrenen Nutzern eine detaillier-

te Sprachaufforderung ausgibt, während bei erfahrenen Nutzern eine verkürzte Sprachaufforderung verwendet. In erfinderischer Weise wird dabei von Seiten des Sprachdialogsystems ein Dialogschritt mit einer verkürzten Sprachaufforderung (Initiierungssignal) initialisiert. Bei Ausbleiben einer Äußerung des Systemnutzer auf die verkürzte Sprachaufforderung wird sodann nach einer bestimmten Zeit (Erkenner-Timeout) eine detaillierte Sprachaufforderung ausgegeben. In vorteilhafter Weise werden dem Systemnutzer also bei jedem Dialogschritt beide Arten von Sprachaufforderungen, eine verkürzte als auch eine detaillierte, zur Verfügung gestellt. Dabei beginnt der Dialogschritt immer mit einer verkürzten Sprachaufforderung, so dass es dem erfahrenen Systemnutzer (Experte) also immer möglich ist die Initiative zu ergreifen; das heißt, dass es diesem immer möglich ist über die Art und Weise des Dialoges zu entscheiden. Dem erfahren Nutzer ist es somit immer Möglich in Bezug auf den Ablauf des Dialoges die Initiative zu ergreifen. Ist auch er sich an einem Punkt des Sprachdialoges unsicher, welche Art oder Weise der Sprachäußerung das Sprachdialogsystem an dieser Stelle erwartet, kann er einfach den Zeitablauf des Erkenner-Timeouts abwarten und erhält sodann eine detaillierte Sprachaufforderung. Bei den nachfolgenden Schritten kann sich der erfahrene Nutzer wieder gleich nach der verkürzten Sprachaufforderung äußern und den Dialog hierdurch beschleunigen.

In Bezug auf die Gestaltung der verkürzten Sprachaufforderung ist es beispielsweise denkbar, diese auf die notwendigste Information oder auf einzelne, die eigentliche Detailinformation besonders bezeichnende Stichworte zu beschränken. Andererseits kann in besonderes vorteilhafter Weise eine Effizienzsteigerung in Bezug auf den Ablauf des Sprachdialoges geschaffen werden, wenn die verkürzte Sprachaufforderung einzig durch eine neutrale, keine spezifische Information enthaltendes Audio-Signal erfolgt; beispielsweise vergleichbar mit der

Sprachaufforderung bei einem Telefonanrufbeantworter, bei welchem der Anrufer aufgefordert wird nach dem Signalton oder dem Piepston (,Beep') zu sprechen.

- Die Effizienz des Verfahrens lässt sich weiter, insbesondere in Bezug auf unerfahrene Systemnutzer, dadurch steigern, dass in einer Speichereinheit protokolliert wird, wie häufig ein Systemnutzer sich erst auf die Ausgabe der detaillierten Sprachaufforderung hin äußert. Äußert sich ein Nutzer wiederholt erst dann, das heißt reagiert er nie oder selten bereits auf die verkürzte Sprachaufforderung, so ist dies ein Hinweis darauf, dass es sich bei diesem um einen unerfahrenen Systemnutzer handeln könnte. In diesem Falle kann in gewinnbringender Weise die Zeitspanne für den Erkenner-Timeout, welcher den Zeitraum zwischen verkürzter und detaillierter Sprachaufforderung definiert, verkürzt werden. Eine sinnvolle Anzahl von für eine Verkürzung des Erkenner-Timeouts notwendigen Wiederholungen könnte auf die Anzahl 3 voreingestellt werden ; d.h. Äußert sich der Systemnutzer dreimal hintereinander erst auf die detaillierte Sprachaufforderung so wird der Erkenner-Timeout verkürzt, beispielsweise halbiert. Hierdurch wäre es auch einem unerfahrenen Systemnutzer möglich den Sprachdialog schneller zu Ziel zu führen. Es ist hierbei denkbar den Erkenner-Timeout dann wieder auf die ursprüngliche Zeitspanne zu setzen, wenn der Systemnutzer in einem der Dialogschritte bereits auf die verkürzte Sprachaufforderung hin reagiert; selbstverständlich ist es auch hier möglich diese Fälle zu protokollieren und den Erkenner-Timeout erst nach mehreren sukzessiven Äußerungen auf eine verkürzte Sprachaufforderung wieder auf den ursprünglichen Wert hin zurückzusetzen.
- In besonderer Weise könnte die Änderung des Erkenner-Timeouts (Verkürzung oder Verlängerung) auch derart gestaltet werden, dass diese sukzessive in mehreren Schritten erfolgt. So könnte die Verkürzung oder nachträgliche Verlängerung des Erkenner-Timeouts weniger abrupt erfolgen. Beträgt die Änderung

für jedes weitere Mal bei dem gleich wie das vorangegangene Mal reagiert wurde beispielsweise 10% der vorherigen Zeitdauer der Erkennen-Timeouts, so würde sich das System annähernd unmerklich an den Systemnutzer anpassen. Das bedeutet, dass

5 für jedes weitere Mal in dem der Systemnutzer sinnvoll erst auf die detaillierte Sprachaufforderung reagierte der Erkennen-Timeout verkürzt würde, und dass er für jedes weitere Mal in welchem er in Folge bereits sinnvoll auf die verkürzte Sprachaufforderung hin antwortete der Erkennen-Timeout in

10 Schritten wieder bis zum ursprünglichen Wert hin erhöht würde. Es wäre hierbei möglich mit der Modifikation des Erkennen-Timeouts bereits nach der ersten Äußerung des Systemnutzers zu beginnen, was die Effizienz des Systems weiter steigern würde.

15 Eine weitere Effizienzsteigerung des Sprachdialogsystems lässt sich dadurch erzielen, dass dieses Barge-In fähig gestaltet wird. Barge-In erlaubt es dem Systemnutzer, die Sprachaufforderungen eines Sprachdialogsystems durch eine eigene sprachliche Eingabe abubrechen. Bei einer solchen

20 sprachlichen Eingabe kann es sich zum einen um die vorzeitige Eingabe der vom System erwarteten Äußerung handeln, oder aber um andere den Sprachdialog beeinflussende Angaben. Durch diese sprachliche Eingabe wird die weitere Ausgabe der Sprachaufforderung unterbrochen. Dies bietet den Vorteil einer effizienteren Interaktion mit dem System, indem der Sprachdialog dadurch beschleunigt wird, dass der Systemnutzer Sprach-

25 aufforderungen unterbrechen und stoppen kann. Hierdurch wird die Möglichkeit geschaffen, dass insbesondere ein erfahrener Systemnutzer, welcher bei einem Dialogschritt Hilfestellung

30 benötigt, die detaillierte Sprachausgabe bereits zu dem Zeitpunkt abbrechen kann, zu welchem er die für die nachfolgende Sprachäußerung notwendigen Hinweise erhalten hat.

In besonders vorteilhafter Weise wird durch die Erfindung ein Sprachdialogsystem geschaffen, welches dynamisch und schnell

auf das aktuelle Bedienverhalten eines Systemnutzers reagieren kann. Ist Systemnutzer mit dem Dialogsystem vertraut, so erlaubt das Verfahren eine effiziente Interaktion, da sofort nach der verkürzten Sprachaufforderung (Initiierungssignal) eine Sprachäußerung getätigt werden kann. Treten dagegen Schwierigkeiten in Bezug auf die zu tätigende Sprachäußerung auf, so reagiert das Sprachdialogsystem entsprechend, indem es eine unterstützende Sprachaufforderung ausgibt. Dabei ist durch der Sprachdialog durch das erfinderische Verfahren gleichzeitig so flexibel ausgelegt, dass falls es zu Schwierigkeiten bei einem der Dialogschritte kommt, dies keine Auswirkungen auf die Reaktionsfähigkeit bei den nachfolgenden Schritten hat. Hat ein Systemnutzer beispielsweise nur deshalb Schwierigkeiten mit der abzugebenden Sprachäußerung, da er gerade abgelenkt war, so wird ihm eine unterstützende Sprachaufforderung präsentiert, auf die er antworten kann. Bei dem nächsten Dialogschritt hat er jedoch wieder die Möglichkeit, sich sofort nach der verkürzten Sprachaufforderung (Initiierungssignal) zu äußern und somit den kürzeren und effizienteren Weg durch den Sprachdialog zu wählen.

DaimlerChrysler AG

Finkele

14.10.2003

Patentansprüche

1. Verfahren für eine nutzeradaptive Dialogführung für ein Sprachdialogsystem,

bei welchem durch das Sprachdialogsystem eine Sprachaufforderung ausgegeben wird,

wobei hierauf das Sprachdialogsystem auf eine Äußerung durch den Systemnutzer wartet, wobei hierzu, um die Nutzeräußerung zu verstehen, ein Spracherkenner aktiviert wird,

wobei das System unerfahrene und erfahrene Nutzer unterscheidet und es unerfahrenen Nutzern eine detaillierte Sprachaufforderung ausgibt, während bei erfahrenen Nutzern eine verkürzte Sprachaufforderung verwendet,

d a d u r c h g e k e n n z e i c h n e t ,
dass von Seiten des Sprachdialogsystems ein Dialogschritt mit einer verkürzten Sprachaufforderung initialisiert wird,

worauf bei Ausbleiben einer Äußerung des Systemnutzer auf die verkürzte Sprachaufforderung nach einer bestimmten Zeit (Erkenner-Timeout) eine detaillierte Sprachaufforderung ausgegeben wird.

2. Verfahren nach Anspruch 1,
d a d u r c h g e k e n n z e i c h n e t ,
dass die verkürzte Sprachaufforderung in Form eines kurzen akustischen Signals (Piepston) erfolgt.
- 5
3. Verfahren nach einem der Ansprüche 1 oder 2,
d a d u r c h g e k e n n z e i c h n e t ,
dass dann, wenn sich der Systemnutzer wiederholt nicht auf die verkürzte Sprachaufforderung äußert, die Zeitspanne für den Erkenner-Timeout, nach welchem eine detaillierte Sprachausgabe erfolgt, verkürzt wird.
- 10
4. Verfahren nach Anspruch 3,
d a d u r c h g e k e n n z e i c h n e t ,
dass die Verkürzung der Zeitspanne für den Erkenner-Timeout mit zunehmender Anzahl von nicht erfolgten Äußerungen auf die verkürzte Sprachaufforderung in mehreren Stufen erfolgt.
- 15
5. Verfahren nach Anspruch 3 oder 4,
d a d u r c h g e k e n n z e i c h n e t ,
dass dann, wenn sich der Systemnutzer bereits auf die verkürzte Sprachaufforderung hin meldet, die Zeitspanne für den Erkenner-Timeout, verlängert wird.
- 20
6. Verfahren nach einem der vorhergehenden Ansprüche,
d a d u r c h g e k e n n z e i c h n e t ,
dass das Sprachdialogsystem so ausgestaltet wird, dass es dem Systemnutzer möglich ist durch vorzeitige Eingabe von Sprachäußerung (Barge-In) die Ausgabe der Sprachaufforderung zu unterbrechen.
- 25
- 30

DaimlerChrysler AG

Finkele

ZusammenfassungNutzeradaptive Dialogunterstützung für Sprachdialogsysteme

Bei Sprachdialogsystemen besteht jedoch generell das Problem, dass man mit dem System unterschiedlich erfahrene Nutzer sprachlich möglichst optimal bedienen muss. Es wird ein Sprachdialogsystem vorgeschlagen, welches unerfahrene und erfahrene Systemnutzer unterscheidet, und daran angepasste Sprachaufforderungen so generiert. Hierbei ist das System in der Lage unerfahrene und erfahrene Nutzer zu unterscheiden, wobei es unerfahrenen Nutzern eine detaillierte Sprachaufforderung ausgibt, während es bei erfahrenen Nutzern eine verkürzte Sprachaufforderung verwendet. In erfinderischer Weise wird dabei von Seiten des Sprachdialogsystems ein Dialogschritt mit einer verkürzten Sprachaufforderung initialisiert. Bei Ausbleiben einer Äußerung des Systemnutzer auf die verkürzte Sprachaufforderung wird sodann nach einer bestimmten Zeit (Erkenner-Timeout) eine detaillierte Sprachaufforderung ausgegeben. So werden also bei jedem Dialogschritt beide Arten von Sprachaufforderungen ausgegeben und dem Systemnutzer zur Auswahl gestellt. Es ist diesem also immer möglich über die Art und Weise des Dialoges zu entscheiden. Dem erfahrenen Nutzer ist es somit immer Möglich in Bezug auf den Ablauf des Dialoges die Initiative zu ergreifen. Ist auch er sich an einem Punkt des Sprachdialoges unsicher, welche Art oder Weise der Sprachäußerung das Sprachdialogsystem an dieser Stelle erwartet, kann er einfach den Zeitablauf des Erkenner-Timeouts abwarten und erhält sodann eine detaillierte Sprachaufforderung.